RR-94-28-ONR

DEALING WITH UNCERTAINTY ABOUT ITEM PARAMETERS: EXPECTED RESPONSE FUNCTIONS

Robert J. Mislevy
Marilyn S. Wingersky
Kathleen M. Sheehan

This research was sponsored in part by the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-88-K-0304
R&T 4421552

Robert J. Mislevy, Principal Investigator

Educational Testing Service
Princeton, NJ

April 1994

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
April, 1994

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Dealing with Uncertainty about Item Parameters: Expected Response Functions

5. FUNDING NUMBERS
G. N00014-88-K-0304
PE. 61153N
PR. RR 04204
TA. RR 04204-01
WU. R&T 4421552

6. AUTHOR(S)
Robert J. Mislevy, Marilyn S. Wingersky, & Kathleen M. Sheehan

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Educational Testing Service
Rosedale Road 03-T
Princeton, NJ 08541

8. PERFORMING ORGANIZATION REPORT NUMBER
RR-94-28-ONR

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Cognitive Sciences
Code 1142CS
Office of Naval Research
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified/Unlimited

12b. DISTRIBUTION CODE
N/A

13. ABSTRACT *(Maximum 200 words)*
It is a common practice in item response theory (IRT) to treat estimates of item parameters, say $\hat{\mathbf{B}}$, as if they were the known, true quantities, **B**. However, ignoring the uncertainty associated with item parameters can lead to biases and over-confidence in subsequent inferences such as ability estimation, especially when item-calibration samples are small. This paper demonstrates how to incorporate uncertainty about **B** with Lewis's "expected response functions" (ERFs), pointwise expected values of item response conditional on examinee proficiency averaged over posterior distributions of item parameters. This paper presents ERFs, outlines procedures for computing them and using them in practical work, and gives an illustration with data from the national Assessment of Educational Progress. Advantages of approximating ERFs response curves with members of familiar parametric families of IRT curves are noted.

14. SUBJECT TERMS
Bayesian estimation, expected response functions, item response theory, multiple imputation, pseudolikelihood estimation

15. NUMBER OF PAGES
35

16. PRICE CODE
N/A

17. SECURITY CLASSIFICATION OF REPORT
Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT
Unclassified

20. LIMITATION OF ABSTRACT
SAR

Dealing with Uncertainty about Item Parameters: Expected Response Functions

Robert J. Mislevy, Marilyn S. Wingersky, & Kathleen M. Sheehan

Educational Testing Service

April, 1994

This work was supported by Contract No. N00014-91-J-4101, R&T 4421573-01, from the Cognitive Science Program, Cognitive and Neural Sciences Division, Office of Naval Research, and by the Program Research Planning Council, Educational Testing Service. We are grateful to Duanli Yan for computing assistance and to Marna Golub-Smith, Charlie Lewis, and Jerry Melican for helpful discussions and comments.

Dealing with Uncertainty about Item Parameters: Expected Response Functions

Abstract

It is a common practice in item response theory (IRT) to treat estimates of item parameters, say $\hat{\mathbf{B}}$, as if they were the known, true quantities, $\mathbf{B}$. However, ignoring the uncertainty associated with item parameters can lead to biases and over-confidence in subsequent inferences such as ability estimation, especially when item-calibration samples are small. This paper demonstrates how to incorporate uncertainty about $\mathbf{B}$ with Lewis's "expected response functions" (ERFs), pointwise expected values of item response conditional on examinee proficiency averaged over posterior distributions of item parameters. This paper presents ERFs, outlines procedures for computing them and using them in practical work, and gives an illustration with data from the National Assessment of Educational Progress. Advantages of approximating ERFs response curves with members of familiar parametric families of IRT curves are noted.

Key words: Bayesian estimation, expected response functions, item response theory, multiple imputation, pseudolikelihood estimation

Introduction

Item response theory (IRT) models posit that an examinee's chances of correctly answering test items depend on an unobservable parameter for that examinee (θ) and for each of the items (β_j, for j=1,...,n). It is common to estimate the item parameters from the response of a "calibration sample" of examinees, then treat the estimates $\hat{\mathbf{B}} = \left(\hat{\beta}_1,\ldots,\hat{\beta}_n\right)$ as if they were true parameter values in subsequent inferences such as estimating examinees' proficiency parameters. Tsutakawa and Johnson (1990) found that ignoring uncertainty about 3-parameter logistic (3PL) item parameters from a calibration sample of 400 led to biased posterior means for θs and understatement of posterior standard deviations by more than 40-percent on the average.

Approaches that take uncertainty about **B** into account include a second-order Taylor series expansion with an asymptotic normal approximation for p(**B**) (Tsutakawa & Soltys, 1988; Tsutakawa & Johnson, 1990), numerical integration over a normal approximation (Jones, Wainer, & Kaplan, 1984), multiple imputation (Mislevy & Yan, 1991), and Gibbs sampling (Albert, 1992). This paper presents approximations based on Lewis's (1985) notion of "expected response functions" (ERFs), pointwise expected values of item response conditional on θ as averaged over posterior distributions of item parameters. (See Mislevy, Sheehan, & Wingersky, 1993, on the use of ERFs in IRT test equating when information about item parameters is limited.)

The following section describes the problem and reviews previous solutions. ERFs and computing approximations are then given. Their use is illustrated with data from the National Assessment of Educational Progress.

Background and Notation

Item Response Theory

This paper confines discusssion to scalar parametric IRT models for dichotomous (right/wrong) test items, but the ideas can be extended to more complex models. Define $F_j(\theta)$, the item response function for Item j, as follows:

$$\mathrm{F}_j(\theta) = \mathrm{Prob}\left(X_j = 1|\theta,\beta_j\right), \tag{1}$$

where X_j is the response to Item j, 1 for right and 0 for wrong, θ is the examinee proficiency parameter, and β_j is the (possibly vector-valued) parameter for Item j. For example, under the 3-parameter logistic (3PL) model,

$$\mathrm{F}_j(\theta) = c_j + (1 - c_j)\Psi\left[1.7a_j(\theta - b_j)\right],$$

where Ψ is the logistic distribution $\Psi(z)=[1+\exp(-z)]^{-1}$ and $\beta_j \equiv (a_j, b_j, c_j)$ (Lord, 1980). The density $\mathrm{p}(x_j|\theta,\beta_j)$ is thus $\mathrm{F}_j(\theta)$ if x_j=1 and 1-$\mathrm{F}_j(\theta)$ if x_j=0. Under the usual IRT assumption of conditional independence, the probability of a vector of responses $x=(x_1,\ldots,x_n)$ to n items is the product over items of terms based on (1):

$$\begin{aligned}\mathrm{p}(x|\theta,\mathbf{B}) &= \prod_{j=1}^{n} \mathrm{p}(x_j|\theta,\beta_j) \\ &= \prod_{j=1}^{n} \mathrm{F}_j(\theta)^{x_j}\left[1-\mathrm{F}_j(\theta)\right]^{1-x_j}.\end{aligned} \tag{2}$$

Equation 2 is the basis for estimating an examinee's θ. Suppose x and $\mathbf{B}$ were known. For maximum likelihood estimation, one finds the value of θ that maximizes (2), namely, the MLE $\hat{\theta}$. The asymptotic variance of the MLE is the inverse of the Fisher information function, which is a sum of contributions over items:

$$\mathrm{Var}^{-1}(\hat{\theta}|\theta,\mathbf{B}) \approx \sum_j \frac{\left[\frac{\partial}{\partial\theta}\mathrm{F}_j(\theta)\right]^2}{\mathrm{F}_j(\theta)\left[1-\mathrm{F}_j(\theta)\right]}. \tag{3}$$

For Bayesian inference, if $\mathrm{p}(\theta)$ represents prior knowledge about an examinee's proficiency before x is observed, then knowledge posterior to the observation is obtained by Bayes theorem as

$$\mathrm{p}(\theta|x,\mathbf{B}) = \frac{\mathrm{p}(x|\theta,\mathbf{B})\,\mathrm{p}(\theta)}{\int \mathrm{p}(x|\theta,\mathbf{B})\,\mathrm{p}(\theta)\,\partial\theta}. \tag{4}$$

The posterior mean and variance are, respectively,

$$\mathrm{E}(\theta|x,\mathbf{B}) = \int \theta\,\mathrm{p}(\theta|x,\mathbf{B})\,\partial\theta \tag{5}$$

and

$$\mathrm{Var}(\theta|\boldsymbol{x},\mathbf{B}) = \int \theta^2\, \mathrm{p}(\theta|\boldsymbol{x},\mathbf{B})\, \partial\theta - \left[\int \theta\, \mathrm{p}(\theta|\boldsymbol{x},\mathbf{B})\, \partial\theta\right]^2. \qquad (6)$$

Uncertainty About Item Parameters

Equations 2 through 6 are written as conditional on **B**. It is common to evaluate such expressions using a point estimate of **B**, or $\hat{\mathbf{B}}$, as obtained for example from the responses $\mathbf{X}_{\text{calib}} = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)$ of a calibration sample of N examinees. For example, the Bayes modal estimate of **B** when p(θ) is known maximizes the posterior distribution for **B**,

$$\mathrm{p}(\mathbf{B}|X_{\text{calib}}) = \mathrm{p}(X_{\text{calib}}|\mathbf{B})\,\mathrm{p}(\mathbf{B}) \propto \prod_{i=1}^{N} \int \mathrm{p}(\boldsymbol{x}_i|\theta,\mathbf{B})\,\mathrm{p}(\theta)\partial\theta\, \mathrm{p}(\mathbf{B}), \qquad (7)$$

where p(**B**) expresses prior knowledge about **B** (e.g., Mislevy, 1986, Tsutakawa, 1984)—perhaps uninformative, perhaps based on items' content or skill requirements, expert judgments, or experience with similar items (Mislevy, Sheehan, & Wingersky, 1993). In large samples, the posterior distribution can be approximated by a multivariate normal distribution with mean $\hat{\mathbf{B}}$ and variance

$$\Sigma_{\mathbf{B}} = \left[\left. \frac{\partial^2 \left[\log \mathrm{p}(\mathbf{B}|X_{\text{calib}})\right]}{\partial\mathbf{B}\partial\mathbf{B}'} \right|_{\mathbf{B}=\hat{\mathbf{B}}} \right]^{-1}. \qquad (8)$$

Values $\hat{\mathbf{B}}$ and $\Sigma_{\mathbf{B}}$ for an approximation could be obtained, for example, as maximum likelihood or Bayesian modal estimates and asymptotic covariance matrix from Mislevy & Bock's (1983) BILOG program, as illustrated in the NAEP example below. In the sequel, we simply use p(**B**) to stand for knowledge about **B** at a given point in time, regardless of its source. Note that p(**B**) need not incorporate independence over items.

As Tsutakawa *et al.* demonstrate, ignoring the uncertainty about **B** (by treating $\hat{\mathbf{B}}$ as **B**) can lead to biases and understated uncertainties in subsequent inferences about θs. Incorporating this kind of uncertainty into analyses is straightforward from a Bayesian perspective: Marginalize with respect to partially-known quantities. For example, the so-called "marginal likelihood function" takes uncertainty about **B** into account in the likelihood function by integrating (2) with respect to p(**B**):

$$
\begin{aligned}
p(x|\theta) &= E_{\mathbf{B}}[p(x|\theta,\mathbf{B})] \\
&= \int p(x|\theta,\mathbf{B})\,p(\mathbf{B})\,\partial\mathbf{B} \\
&= \int \prod_{j=1}^{n} p(x_j|\theta,\beta_j)\,p(\mathbf{B})\,\partial\mathbf{B} \\
&= \int \prod_{j=1}^{n} F_j(\theta)^{x_j}\left[1-F_j(\theta)\right]^{1-x_j} p(\mathbf{B})\,\partial\mathbf{B},
\end{aligned} \tag{9}
$$

effectively the average of (2) over all possible values of **B**, each weighted by its probability given the information from the calibration sample. More generally, if G(**B**) is any expression involving item parameters, then

$$
E_{\mathbf{B}}[G(\mathbf{B})] = \int G(\mathbf{B})\,p(\mathbf{B})\,\partial\mathbf{B}. \tag{10}
$$

Alternative Approaches

Closed-form solutions of (10) are not generally forthcoming in IRT. Before introducing expected response functions, we briefly review three alternatives: a second-order analytic approximation, multiple imputation, and Gibbs sampling. The discussion of multiple imputation is more detailed, because the ERF approximation shares intermediary steps with multiple imputation and the NAEP example compares numerical results from the two approaches.

Tsutakawa's <u>second-order expansion</u> uses an approximation due to Lindley (1980):

$$
E_{\mathbf{B}}[G(\mathbf{B})] \approx G(\hat{\mathbf{B}}) + \tfrac{1}{2}\sum_{r,s} G_{rs}\Sigma_{rs}, \tag{11}
$$

where G_{rs} is the r,s^{th} element of $\partial^2[G(\mathbf{B})]/\partial\mathbf{B}\partial\mathbf{B}'$ and Σ_{rs} is the r,s^{th} element of $\Sigma_{\mathbf{B}}$, with r and s indexing elements of **B**. When calculating an examinee's posterior mean (5), for example, G(**B**) is $\int \theta\, p(\theta|x,\mathbf{B})\,\partial\theta$. Because such approximations would be exact if p(**B**) were MVN($\hat{\mathbf{B}}$,$\Sigma_{\mathbf{B}}$), their performance in (10) depends on the accuracy of the asymptotic normal approximation to p(**B**)—which is often satisfactory in practice since even the usual first-order approximation G($\hat{\mathbf{B}}$) suffices when the calibration sample is large and p(**B**|*X*) is concentrated around $\hat{\mathbf{B}}$. An impediment to using (11) in practical work is that derivatives must be calculated for each function G to which it is applied.

Albert (1992) employed Gibbs sampling (Gelfand & Smith, 1990) to obtain a discrete approximation to the joint posterior distribution of **B** and the vector of examinee abilities $\boldsymbol{\Theta}$ under the 2-parameter normal (2PN) IRT model. From vectors $\mathbf{B}^{(t)}$ and $\boldsymbol{\Theta}^{(t)}$ that approximate **B** and $\boldsymbol{\Theta}$, one obtains a subsequent approximation by drawing $\mathbf{B}^{(t+1)}$ from $p\left(\mathbf{B} \mid \boldsymbol{\Theta} = \boldsymbol{\Theta}^{(t)}, \boldsymbol{X}\right)$, then drawing $\boldsymbol{\Theta}^{(t+1)}$ from $p\left(\boldsymbol{\Theta} \mid \mathbf{B} = \mathbf{B}^{(t+1)}, \boldsymbol{X}\right)$. From initial approximations, repeated cycles achieve (under regularity conditions) a stochastic convergence such that a $(\boldsymbol{\Theta},\mathbf{B})$ draw obtained in this manner is essentially a draw from the correct posterior $p(\boldsymbol{\Theta}, \mathbf{B} \mid \boldsymbol{X})$. Widely spaced draws from a sequence which has attained convergence (or, better still, from separate sequences initiated from different starting points; see Gelman & Rubin, 1992) are essentially independent draws from $p(\boldsymbol{\Theta}, \mathbf{B} \mid \boldsymbol{X})$. Evaluating any function $G(\boldsymbol{\Theta},\mathbf{B})$ of the parameters with respect to each of these draws constitutes a discrete approximation of its posterior distribution. (This last idea will be illustrated below with multiple imputation.) In particular, the discrete approximation of p(**B**) can serve as a basis for calculating expected response functions. Gibbs sampling is much more computationally intensive than the other approximations described in this paper.

Multiple imputation, introduced by Rubin (1987) to handle missing responses in sample surveys, creates pseudo datasets with draws from the posterior distributions of missing data, and combines the results of standard analyses of pseudo data sets so as to incorporate the uncertainty that missingness engenders. **B** plays the role of missing data in the problem of imperfect knowledge about item parameters (Mislevy & Yan, 1991). Suppose that if **B** were known, we could calculate the posterior mean and variance of $G(\mathbf{B})$, say, $\overline{G}(\mathbf{B})$ and $V(\mathbf{B})$. An example again would be the posterior mean and variance for an examinee's θ via (5) and (6). The steps for multiple-imputation approximations of the posterior mean and variance that take uncertainty about **B** into account, say, $\overline{\overline{G}}$ and $\overline{\overline{V}}$, are outlined below. The reader is referred to Rubin (1987) for theoretical justification.

1. Obtain the posterior distribution for **B**, p(**B**) (e.g., the multivariate normal approximation $MVN(\hat{\mathbf{B}},\Sigma_{\mathbf{B}})$ used in the following NAEP example).

2. Draw K item parameter vectors from p(**B**), say $\mathbf{B}_k$ for $k=1,\ldots,K$.

3. For each k, calculate the posterior mean and variance conditional on $\mathbf{B}=\mathbf{B}_k$, denoted $\overline{G}(\boldsymbol{B}_k)$ and $V(\mathbf{B}_k)$.

4. The posterior mean for G, accounting for uncertainty about **B**, is approximated by the average of the K conditional posterior means:

$$\overline{\overline{G}} = K^{-1}\sum_k \overline{G}(\mathbf{B}_k). \tag{12}$$

5. The posterior variance for G, accounting for uncertainty about **B**, is approximated by the sum of two terms:

$$\overline{\overline{V}} = U + \tfrac{K+1}{K}\overline{V}, \tag{13}$$

where the first,

$$U = K^{-1}\sum_k V(\mathbf{B}_k),$$

approximates the variance that would exist even if **B** were known with certainty, and the second,

$$\overline{V} = (K-1)^{-1}\sum_k \left[\overline{G}(\mathbf{B}_k) - \overline{\overline{G}}\right]^2$$

quantifies additional uncertainty due to not knowing **B**.

Example: Data from NAEP

We shall use a running example with data from the National Assessment of Educational Progress (NAEP): responses to 19 items from 100 8- and 13-year old students who participated in the 1986 and 1988 mathematics trend assessment. Table 1 gives descriptive statistics and Bayesian posterior modal estimates $\hat{B} = (\hat{a}, \hat{b}, \hat{c})$ obtained with Mislevy and Bock's (1983) BILOG computer program. Table 2 gives the accompanying approximation of the posterior covariance matrix $\Sigma_{\mathbf{B}}$. Covariances among the three parameters for the same item can be quite high, but relationships among parameters for different items are uniformly much lower.

[[Tables 1 & 2 about here]]

A practical problem in applying multiple imputations is to determine the value of K that provides the desired accuracy, which may differ with the target G. In the NAEP example, Mislevy and Yan (1991) calculated examinees' posterior means and variances with K=10, 100, and 1000. K=10 proved stable for estimating posterior means, but not for posterior variances, which were stable with K=100. Results for K=100 and K=1000 were indistinguishable. We use the K=100 results below as a baseline comparison for

corresponding estimates calculated with ERFs. The dotted lines in Figure 1 illustrate the item response functions for four items from the NAEP example that correspond to 100 draws of **B**. (The solid and dashed lines will be discussed below). These graphs depict the nature and magnitude of uncertainty about item response functions, but not the mild co-relationship among the curves induced by the nonzero inter-item covariances.

[[Figure 1 about here]]

Expected Response Functions

Definition

In dichotomous IRT models, the expected value of a correct response to Item j given θ and **B** is $F_j(\theta) \equiv P(x_j=1|\theta,\beta_j)$. If β_j is only partially known, through p(**B**), the probability of a correct response conditional on θ but marginal with respect to **B** can be written as

$$\begin{aligned} F_j^*(\theta) &\equiv E_{\beta_j}\left[F_j(\theta)\right] \\ &= \int P\left(X_j = 1|\theta,\beta_j\right)p(\mathbf{B})\partial\mathbf{B} \\ &= \int P\left(X_j = 1|\theta,\beta_j\right)p\left(\beta_j\right)\partial\beta_j, \end{aligned} \tag{14}$$

an "expected response function" that gives the probability of correct response conditional on θ taking into account uncertainty about **B** (Lewis, 1985).

Even though F_j^* is the expected value of a correct response at each value of θ, it is *not* the same as $F_j(\theta)$ evaluated with the expected value of β_j. This can be seen in Figure 1, which shows expected response functions (dashed lines) for the four items from the NAEP example, along with the curves that correspond to $F_j(\theta)$ as evaluated with the point estimate $\hat{\beta}_j$ (solid lines). In particular, the ERF is generally flatter.

The shape of F_j^* depends on the shape of F_j and the character of $p(\beta_j)$. In general, F_j^* and F_j will not be of the same functional form. Lewis (1985) shows that if F_j were 2PN and $p(\beta_j) \equiv p(a_j,b_j)$ were bivariate normal, then F_j^* would be a 2-parameter ogive with a Student's t shape. Its location parameter, b_j^*, would have the same value as the Bayes mean estimate for b_j, or $\hat{b}_j$, but its slope parameter, a_j^*, would be attenuated from the

Bayes mean estimate for a_j. A simpler result is obtained if a_j is known with certainty *a priori*. If p(b_j) is N($\hat{b}_j$,σ_j), then F_j^* is also 2PN, with $b_j^* = \hat{b}_j$ and

$$a_j^* = \left(a_j^{-2} + \sigma_j^2\right)^{-\frac{1}{2}}.$$

Approximation with ERFs

ERFs serve as a potential basis for taking uncertainty about **B** into account, by replacing occurrences of F_js with F_j^*s in functions of interest G(**B**). As examples, consider the following:

<u>Likelihood estimation of θ</u> proceeds by maximizing an ERF-based analogue of the likelihood, namely

$$p^*(x|\theta) \equiv \prod_{j=1}^{n} F_j^*(\theta)^{x_j}\left[1 - F_j^*(\theta)\right]^{1-x_j}. \tag{15}$$

One way to justify maximizing $p^*(x|\theta)$ is to view it as an approximation of the marginal likelihood:

$$\begin{aligned}
p(x|\theta) &= E_{\mathbf{B}}\left[p(x|\theta,\mathbf{B})\right] \\
&= \int \prod_{j=1}^{n} F_j(\theta)^{x_j}\left[1 - F_j(\theta)\right]^{1-x_j} p(\mathbf{B})\,\partial\mathbf{B} \\
&= \int\cdots\int \prod_{j=1}^{n} F_j(\theta)^{x_j}\left[1 - F_j(\theta)\right]^{1-x_j} p\left(\beta_j|\beta_{j-1},\ldots,\beta_1\right)\partial\beta_j \\
&\approx \int\cdots\int \prod_{j=1}^{n} F_j(\theta)^{x_j}\left[1 - F_j(\theta)\right]^{1-x_j} p\left(\beta_j\right)\partial\beta_j \\
&= \prod_{j=1}^{n} \int\cdots\int F_j(\theta)^{x_j}\left[1 - F_j(\theta)\right]^{1-x_j} p\left(\beta_j\right)\partial\beta_j \\
&= \prod_{j=1}^{n} F_j^*(\theta)^{x_j}\left[1 - F_j^*(\theta)\right]^{1-x_j} \\
&\equiv p^*(x|\theta).
\end{aligned}$$

The step in which the approximation occurs replaces each $p\left(\beta_j|\beta_{j-1},\ldots,\beta_1\right)$ with $p\left(\beta_j\right)$. Thus, if the information about items is independent—that is, p(**B**)=$\prod$ p(β_j)—the result is exact. Likelihood and Bayesian inferences about θ that take uncertainty about **B**

into account exhibit in this case the same conditional independence form as when item parameters are known. In particular, applying standard procedures for known item response functions to obtain MLEs and asymptotic variances (3), but with F_j^*s in place of F_js, gives the correct results. Independent posteriors for items can be assured or closely approximated by coupling special item-calibration sampling designs and test construction designs; the idea is for the items appearing in a test, the sets of examinees in the calibration sample responding to each of them were completely or nearly disjoint. For example, randomly equivalent calibration samples of examinees can be administered disjoint blocks of items, and operational test forms can be built with items from different blocks.

A second justification applies even if p(**B**) is not independent over items. Although the dependencies among items are ignored, (15) is an example of what Arnold and Strauss (1991) call a "pseudo-likelihood" (see Appendix); under regularity conditions on the F_j^*s, its maximum is a consistent estimator of θ. Thus likelihood point estimates of θ based on (15) tend to have the correct central tendency. Applying the standard MLE variance formula (3) with F_j^*s tends to give too optimistic of an impression of the uncertainty about θs, however. But if the dependencies among items are small—and they tend toward zero in long tests (Mislevy & Sheehan, 1989)—the degree to which this value understates uncertainty will also be small.

<u>Bayesian inference about θ</u> can employ the above approximation $p^*(x|\theta)$ for likelihoods. The posterior distribution for θ is thus approximated as

$$p^*(\theta|x) = \frac{p^*(x|\theta)\, p(\theta)}{\int p^*(x|\theta)\, p(\theta)\, \partial\theta} \quad ,$$

and the posterior mean and variance are approximated as

$$\begin{aligned} E(\theta|x) &= \int\int \theta\, p(\theta|x,\mathbf{B})\, \partial\theta\partial\mathbf{B} \\ &\approx \int \theta\, p^*(\theta|x)\, \partial\theta \end{aligned} \tag{16}$$

and

$$\begin{aligned} Var(\theta|x) &= \int\int \theta^2\, p(\theta|x,\mathbf{B})\, \partial\theta - \left[\int \theta\, p(\theta|x,\mathbf{B})\, \partial\theta\right]^2 \partial\mathbf{B} \\ &\approx \int \theta^2\, p^*(\theta|x)\, \partial\theta - \left[\int \theta\, p^*(\theta|x)\, \partial\theta\right]^2. \end{aligned} \tag{17}$$

Again the approximations are exact if p(**B**) is independent over items, and indicators of uncertainty tend to be optimistic to the extent that dependencies among items are nonnegligible. Some numerical results on this point appear in the NAEP example.

The test characteristic function is the expected number-correct score on a test of n items as a function of θ. Mislevy, Sheehan, & Wingersky (1993) obtained test characteristic functions with ERFs, in order to equate tests with sparse item-calibration data. IRT true-score equating determines number-right (or formula) scores on different tests that correspond to the same values of θ (Lord, 1980). The expected number-right score on Test A for an examinee with proficiency θ is obtained as

$$\tau_A(\theta) = \sum_{j \in T_A} p(x = 1 | \theta, \beta_j) = \sum_{j \in T_A} F_j(\theta), \tag{18}$$

where T_A is the set of indices of items that appear in Test A. The expected score on Test B, $\tau_B(\theta)$, is defined analogously. A score on Test A and a score on Test B are "true-score equated" if they are the respective expected scores of the same value of θ.

When knowledge about **B** is imperfect, one must equate scores that are expectations conditional on θ but marginal with respect to p(**B**), rather than expected scores conditional on θ and **B**. The expected true score on Test A given θ under these circumstances is thus

$$\tau_A^*(\theta) = E_B[\tau_A(\theta)] = \sum_{j \in T_A} \int p(x = 1 | \theta, \beta_j) p(\beta_j) \partial\beta_j = \sum_{j \in T_A} F_j^*(\theta). \tag{19}$$

This is simply the sum of the probabilities of correct response item by item, whether or not p(**B**) is independent over items. A score on Test A and a score on Test B are "*expected* true-score equated" if they are the respective expected scores of the same value of θ, as defined by (19). Because only expected scores are needed for this equating method, the expected test characteristic curves obtained in (19) are correct whether or not the posteriors for individual items are independent.

Computing Approximations

As noted above, closed-form solutions for F_j^* are not generally available. This section describes how to use multiple-imputations or Gibbs-sampler discrete estimates of p(β_j) to estimate F_j^* point by point across a grid of θ values for each item. Because only p(β_j) is involved for Item j, not the posteriors for other items, this process can be carried

out independently over items. Subsequent inferences about θ can be drawn using these points in a discrete approximation of the θ distribution and the response curve, or a smooth curve can be fit to the probabilities thus obtained.

There are operational advantages to using the closest curve from a familiar family to approximate F_j^*—for example, the closest 3PL curve in applications based on the 3PL model, or the closest 2PL model in applications based on the 1PL or 2PL. Let F_j^{**} denote such an approximation. This expedient makes it possible to use standard off-the-shelf software designed for popular parametric IRT models to estimate examinee scores, construct tests, or draw equating lines. If additional information about item parameters becomes available over time, as might occur as examinee responses are acquired over time in operational testing, it can be incorporated into the system by merely updating item parameter values under the same model. If the IRT model were correct and the response function were stable over time, the sequence of expected response curves would converge toward the closest member of the family to the true curve—to the true curve itself, if it were a member of the family.

We now describe the operational procedures we have used for applied work with ERFs. The expected response function for a particular item, F_j^*, is approximated as follows:

1. Obtain an estimate of the posterior distribution $p(\beta_j)$. As noted above, this is usually based on a calibration sample of examinee responses—say, $MVN\left(\hat{\beta}_j, \Sigma_{\hat{\beta}_j}\right)$ with parameter estimates from BILOG— but it may also be based partly or wholly on collateral information about items such as content specifications and cognitive processing requirements (Mislevy, Sheehan, & Wingersky, 1993).

2. Specify a grid of M theta values across the ability range of interest. Let Θ_m denote the m^{th} grid point.

3. Draw K item parameter vectors from $p(\beta_j)$. Let $\beta_j^{(k)}$ be the k^{th} such draw.

4. For each of the K sets of item parameters, determine $P_{jm}^{(k)}$, the probability of a correct response to Item j at Θ_m, where $P_{jm}^{(k)} = p\left(x_j = 1 | \theta = \Theta_m, \beta_j = \beta_j^{(k)}\right)$.

5. Compute the expectation at each point Θ_m by averaging the probabilities obtained in Step 4:

$$F_j^*(\Theta_m) = K^{-1}\sum_{k=1}^{K} P_{jm}^{(k)}.$$

We refer to the collection of points $\{(\Theta_m, F_j^*(\Theta_m)): m=1,\ldots,M\}$ as the "nonparametric" expected response function because it does not assume any particular parametric form.

For applied work, it may be convenient to approximate the nonparametric ERF with a continuous approximation F_j^{**}, say a spline or a close-fitting 2PL or 3PL curve. The use of a 3PL will be illustrated below. Maximum likelihood estimates for the 3PL item parameters $\beta_j^{**} = \left(a_j^{**}, b_j^{**}, c_j^{**}\right)$ that best approximate F_j^* are found by maximizing

$$\prod_{m=1}^{M}\left\{F_j^{**}\left(\Theta_m;\beta_j^{**}\right)^{F_j^*(\Theta_m)}\left[1 - F_j^{**}\left(\Theta_m;\beta_j^{**}\right)\right]^{1-F_j^*(\Theta_m)}\right\}^{W_m} \tag{20}$$

over the M-point theta grid, where W_m is a weight that specifies the relative importance of fitting F^{**} at Θ_m. For example, weights may be selected to simulate a rectangular distribution of examinees or a normal distribution of examinees. The maximum may be obtained iteratively by using Newton's method to obtain successive corrections to the parameter estimates. We refer to the solution as a "fitted" expected response function.

Example (continued)

The BILOG calibration of the 19 previously-described NAEP items with 100 examinees provided the posterior mode estimates $\left(\hat{a}_j, \hat{b}_j, \hat{c}_j\right)$ and the corresponding large-sample approximation of the covariance matrix discussed above. Due to range restrictions on the a's and c's, we worked with a multivariate normal (MVN) approximation for the posterior of $\beta_j = (\log(a_j), b_j, \text{logit}(c_j))$, where $\text{logit}(c_j)=\log[c_j/(1-c_j)]$. $p(\beta_j)$ was thus approximated as MVN with mean vector $\hat{\beta}_j = (\log(\hat{a}_j), \hat{b}_j, \text{logit}(\hat{c}_j))$ and covariance matrix Σ_{β_j} obtained through the delta method from the covariance matrix for the untransformed parameters. Nonparametric and fitted 3PL ERFs were calculated for each item. Figure 2 presents results for the four items which previously appeared in Figure 1. The nonparametric ERFs were obtained using 100 draws from $p(\beta_j)$ and a grid of 31 evenly-spaced Θ values ranging from -3 to +3 in steps of .2. The fitted curves employed a

standard normal weighting function over the same range. The item response functions that correspond to $F_j(\theta)$ evaluated with the point estimate $\hat{\beta}_j$ are also plotted for comparison. These curves are noticeably steeper than the two expected response curves. Thus, one effect of ignoring uncertainty about item parameters is a tendency to inflate belief about the discriminating power of an item.

[[Figure 2 about here]]

For most of the 19 items, the 3PL approximation captured the nonparametric approximation quite well. The only discrepancies encountered were for items with fairly high a's, such as Item 19. For these highly discriminating items, the fitted curves tended to be slightly flatter than the nonparametric curves. The discrepancies were slightly more pronounced when the ERFs were recalculated with a rectangular weighting function, indicating that they are related to the inability of the 3PL form to capture the pattern of curvature in the tails of the theta range.

Figure 3 presents a comparison of results regarding Bayesian inference about θ for a sample of 100 students. The plots show posterior means and associated posterior standard deviations (PSDs) calculated using point estimates of the item parameters, nonparametric ERFs, and fitted ERFs. In each case, the multiple imputation solution (Equations 12 and 13) is employed as a standard of evaluation, as it is nonparametric and accounts for dependencies among the parameters of different items. As can be seen, the various methods for handling uncertainty about β_j have had negligible effect on the calculation of posterior means. However, the effect on the associated PSDs is quite pronounced. As would be expected, the practice of using point estimates of item parameters as if they were known true values seriously understates the uncertainty associated with examinees' θs. This effect is less pronounced when ERFs are used. Table 3 presents average PSDs calculated for the multiple imputation approach, the nonparametric and fitted ERFs, and the point estimates. In this example, the PSD of a typical examinee's θ, when calculated using point estimates of the item parameters, was understated by about 10%. This can be attributed to ignoring uncertainty about **B** altogether. For the nonparametric and fitted ERFs the understatement was only 3.6% and 3.9% respectively. This is obtained by incorporating uncertainty about **B** item by item, but ignoring dependencies across items. In terms of variance, about 60% of the typically-ignored variance was accounted for in this example through the use of ERFs.

[[Figure 3 about here]]

[[Table 3 about here]]

Conclusion

As increasingly ambitious applications push item response theory closer to the boundaries of its applicability, increasingly strenuous efforts are required to deal with issues of uncertainty, both as to model fit and knowledge of parameters within the model. This paper addresses a problem of the latter type, namely, dealing with uncertainty about item parameters. Fortunately, statisticians' recent interest in numerical and Bayesian approaches to such problems provide a variety of tools, each with their own strengths and weaknesses to be matched with the purposes and characteristics of applications. Expected response functions (ERFs) account for uncertainty that is usually ignored in a way that allow us to employ familiar formulas for known item response functions—even to apply the same formulas but with attenuated parameter estimates. This would be especially convenient in item-banking and adaptive-testing applications, in which tests are assembled from collections of pre-calibrated items. Uncertainty about item parameters (under the assumed model!) would be *implicit in the parameter* estimates available at a given point in time, no additional steps would be required at the point of calculating scores for individual examinees, and improved knowledge about item parameters would merely require updating a file of ERF parameters.

References

Albert, J.H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics, 17,* 251-269.

Arnold, B. C. and Strauss, D. (1991) Pseudolikelihood estimation: Some examples. *Sankhya B, 53,* 233-243.

Gelfand, A.E., & Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85,* 398-409.

Gelman, A., & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 4,* 457-511.

Jones, D.H., Wainer, H., & Kaplan, B. (1984). Estimating ability with three item response models when the models are wrong and their parameters are inaccurate. ETS Research Report RR-84-26. Princeton, NJ: Educational Testing Service.

Lewis, C. (1985). Estimating individual abilities with imperfectly known item response functions. Paper presented at the Annual Meeting of the Psychometric Society, Nashville TN, June, 1985.

Lindley, D.V. (1980). Approximate Bayesian methods. *Trabajos Estadistica, 31,* 223-237.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Mislevy, R.J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51,* 177-196.

Mislevy, R.J., & Bock, R.D. (1983). *BILOG: Item analysis and test scoring with binary logistic models* [computer program]. Mooresville, IN: Scientific Software, Inc.

Mislevy, R.J., & Sheehan, K.M. (1989). Information matrices in latent-variable models. *Journal of Educational Statistics, 14,* 335-350.

Mislevy, R.J., Sheehan, K.M., & Wingersky, M.S. (1993). How to equate tests with little or no data. *Journal of Educational Measurement, 30,* 55-78.

Mislevy, R.J., & Yan, D. (1991, June). Dealing with uncertainty about item parameters: Multiple imputations and SIR. Presented at the annual meeting of the Psychometric Society, Princeton, NJ.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Tsutakawa, R.K. (1984). Estimation of two-parameter logistic item response curves. *Journal of Educational Statistics, 9*, 263-276.

Tsutakawa, R.K., & Johnson, J. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika, 55*, 371-390.

Tsutakawa,R.K., & Soltys, M.J. (1988). Approximation for Bayesian ability estimation. *Journal of Educational Statistics, 13*, 117-130.

Sheehan, K.M., & Mislevy, R.J. (1988). Some consequences of the uncertainty in IRT linking procedures. *Research Report 88-38-ONR*. Princeton, NJ: Educational Testing Service.

Table 1

Statistics and Point Estimates of Item Parameters ($\hat{a}$,$\hat{b}$,$\hat{c}$) for 19 NAEP Mathematics Items

Item	% Correct	r-bis	$\hat{a}$	$\hat{b}$	$\hat{c}$
1	.78	.35	.39	-1.59	.20
2	.92	.63	.90	-1.98	.20
3	.78	.45	.55	-1.23	.19
4	.85	.45	.77	-1.45	.20
5	.79	.45	.63	-1.17	.20
6	.91	.47	.54	-2.60	.20
7	.65	.65	1.20	-.26	.17
8	.86	.64	.99	-1.37	.18
9	.72	.62	1.22	-.50	.19
10	.67	.61	1.27	-.26	.20
11	.48	.56	1.96	.53	.23
12	.77	.44	.60	-1.06	.20
13	.85	.59	.95	-1.30	.19
14	.51	.69	1.89	.20	.15
15	.55	.49	.86	.19	.18
16	.43	.41	.65	.81	.16
17	.30	.56	1.10	1.04	.12
18	.53	.44	2.59	.56	.30
19	.21	.63	3.03	1.09	.10

Table 2

Variances and Covariances of Item Parameter Estimates for 19 NAEP Mathematics Items

Item	Var(a)	Cov(a,b)	Var(b)	Cov(a,c)	Cov(b,c)	Var(c)
1	.059	.233	1.212	.001	.027	.008
2	.435	.632	1.086	.003	.012	.008
3	.069	.141	.509	.002	.019	.008
4	.264	.320	.536	.004	.016	.008
5	.119	.174	.401	.004	.020	.008
6	.078	.325	1.673	.001	.017	.008
7	.300	.055	.073	.011	.010	.006
8	.208	.179	.251	.003	.011	.007
9	.259	.094	.108	.010	.012	.007
10	.339	.074	.077	.016	.012	.007
11	2.513	.056	.058	.053	.008	.006
12	.114	.181	.527	.004	.021	.008
13	.280	.261	.354	.004	.012	.007
14	1.519	.073	.041	.034	.007	.004
15	.203	.037	.132	.011	.015	.007
16	.118	-.043	.201	.009	.014	.006
17	.366	-.075	.104	.012	.005	.003
18	10.944	.232	.058	.129	.009	.008
19	11.626	-.210	.051	.042	.002	.002

Table 3

Average Posterior Variances and Standard Deviations for a Sample of 100 Examinees

Estimation Method	Average Posterior Variance	Average Posterior S.D.	% Decrease
Multiple Imputation	0.2151	.4585	---
Nonparametric ERF	0.1995	.4418	3.6
Fitted ERF	0.1977	.4406	3.9
Point Estimates	0.1743	.4113	10.3

Figure Captions

Figure 1. 100 Draws from Item Parameter Posterior Distributions for Four Items.

Figure 2. Item Response Functions for the Four Items.

Figure 3. Scatterplots of Posterior Means and Standard Deviations for 100 Examinees.

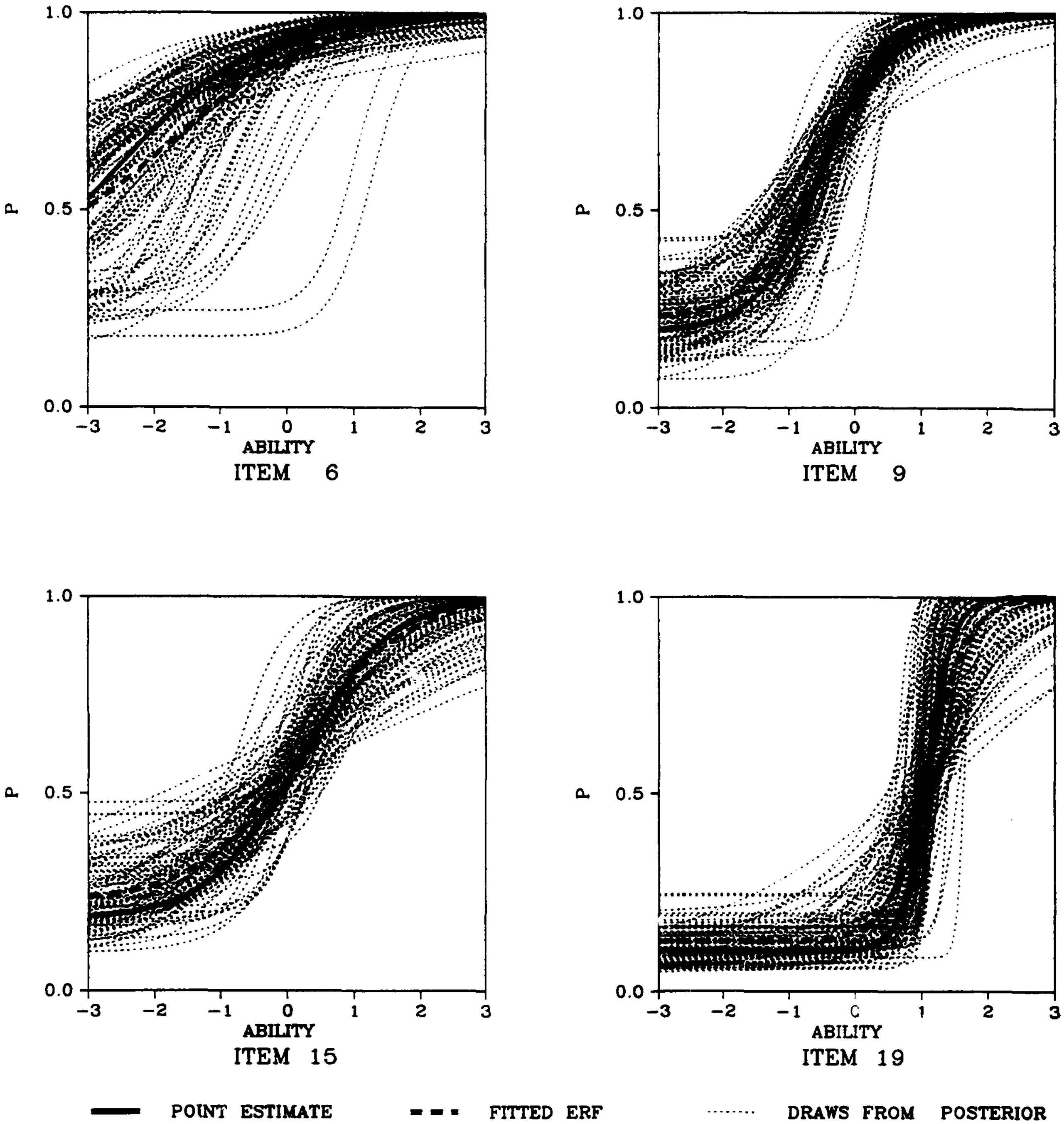

Figure 1

100 Draws from Item Parameter Posterior Distributions for Four Items

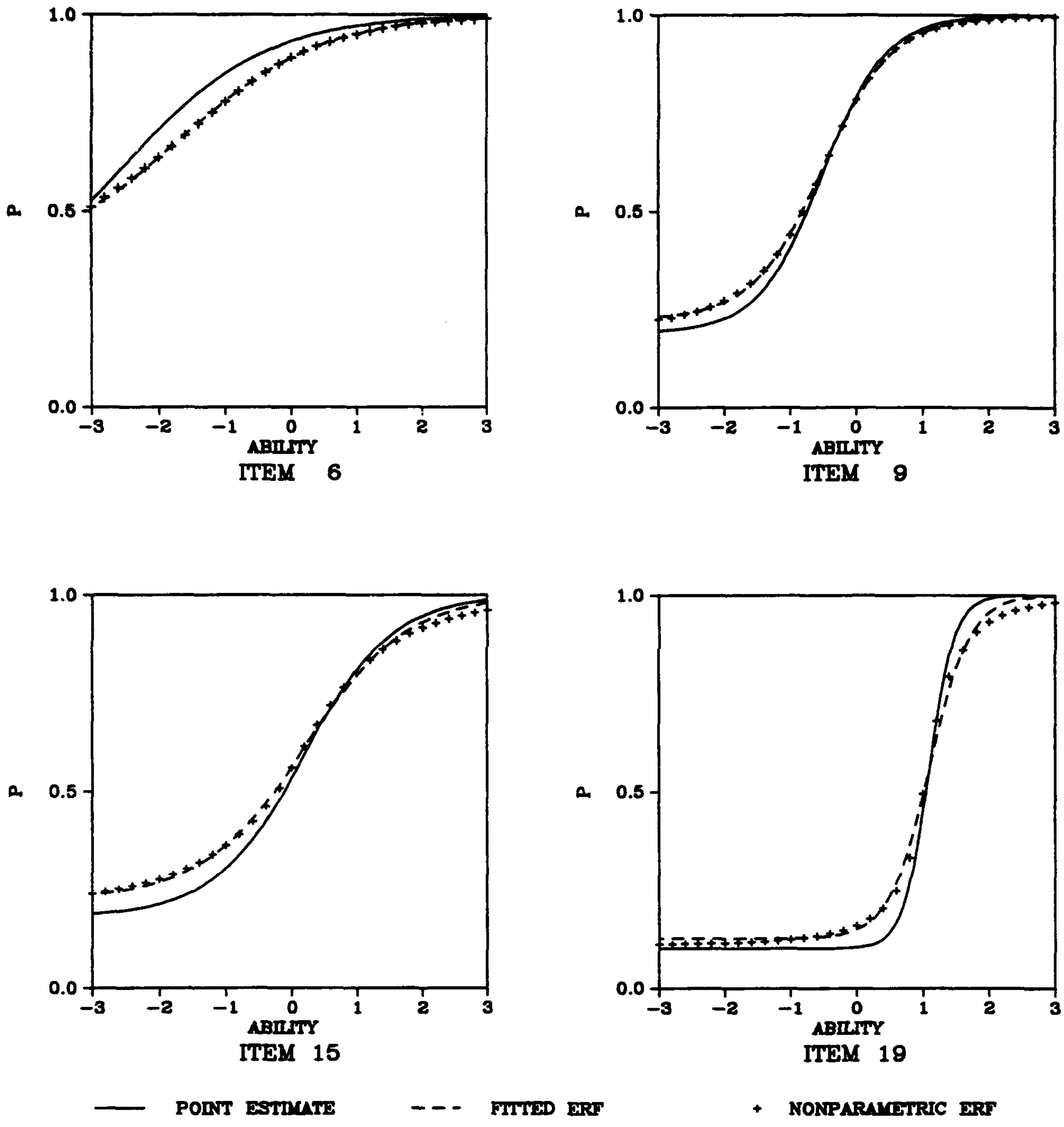

Figure 2
Item Response Functions for the Four Items

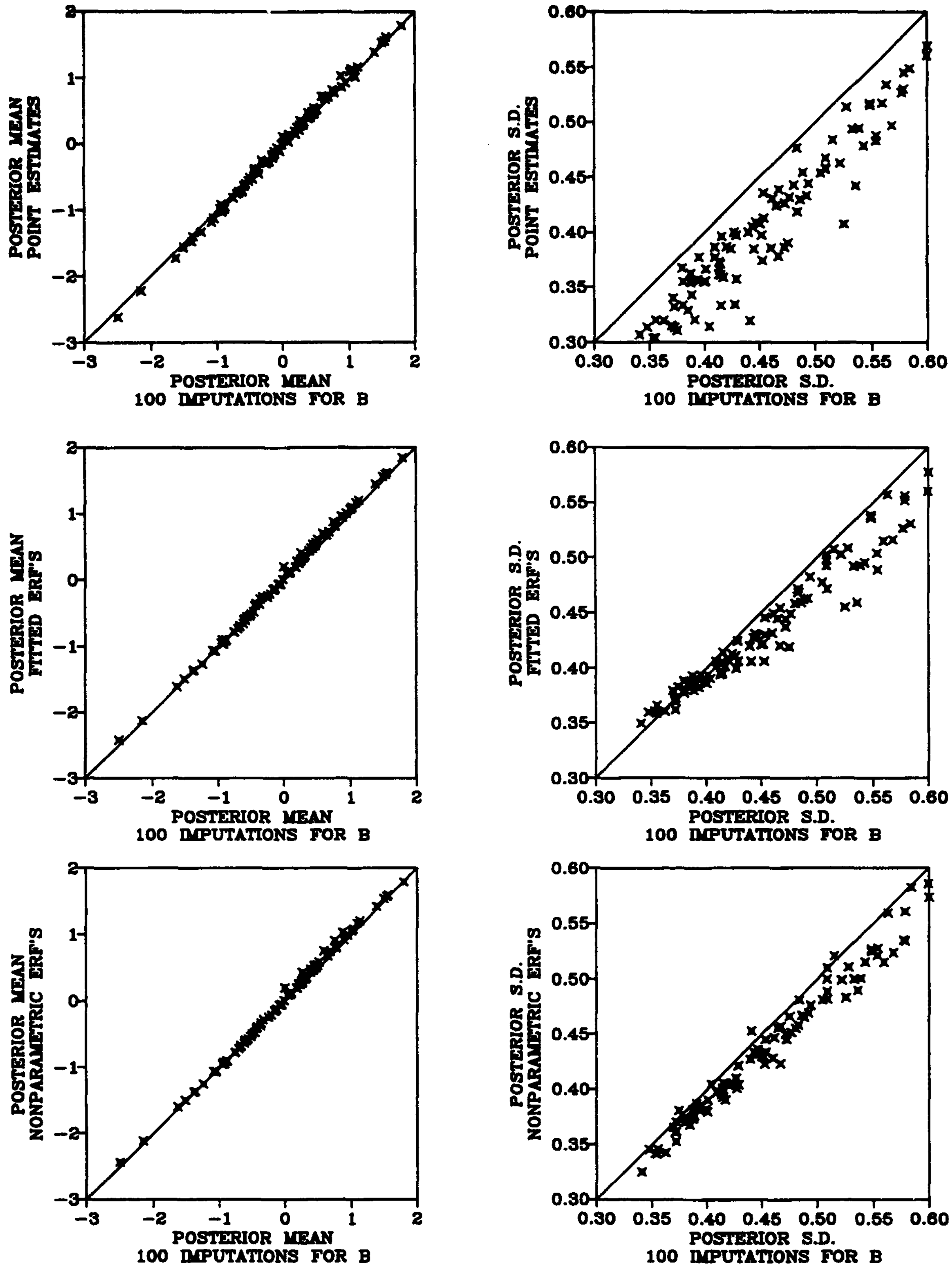

Figure 3
Scatterplots of Posterior Means and Standard Deviations for 100 Examinees

Appendix A

Pseudolikelihood Estimation of θ from Marginalized Likelihoods

The first section below paraphrases Arnold and Strauss's (1991; denoted AS below) framework and results on pseudolikelihood estimation. The reader is referred to AS for regularity conditions, proofs, and examples. The second section shows how this framework accommodates likelihood estimation of θ using the product of expected response curves.

Pseudolikelihood Estimation

Let $(\mathbf{X}_1,\ldots,\mathbf{X}_N)$ represent N iid n-dimensional observations with common joint density $f(x;\theta)$ where θ is an element of a p-dimensional parameter domain Θ. Denote by S the set of all n-dimensional vectors consisting of 0's and 1's, with at least one 1. For a particular s in S, the random vector $\mathbf{X}_i^{(s)}$ contains the coordinates X_{ij} of $\mathbf{X}_i$ for which $s_j=1$. For example, if $\mathbf{X}_i=(X_{i1}, X_{i2}, X_{i3})$ and $s=(1,0,1)$, then $\mathbf{X}_i^{(s)}=(X_{i1}, X_{i3})$. The density of $\mathbf{X}_i^{(s)}$ will be denoted $f_s(x^{(s)};\theta)$, although it may depend on only some of the components of θ. Let $\delta=\{\delta_s : s\in S\}$ be a vector of 2^n-1 real numbers, not all zero, corresponding to the elements of S. The *pseudolikelihood* PL(δ,θ) of the data is defined by

$$\mathrm{PL}(\delta,\theta)=\prod_{s\in S}\left[\prod_{i=1}^{N} f_s\left(x_i^{(s)};\theta\right)\right]^{\delta_s}. \tag{A1}$$

Equivalently, in terms of logarithms,

$$\log \mathrm{PL}(\delta,\theta)=\sum_{s\in S}\delta_s\sum_{i=1}^{N}\log f_s\left(x_i^{(s)};\theta\right).$$

A pseudolikelihood(δ) estimate of θ is a value of θ that maximizes (A1). Under regularity conditions, (A1) can be maximized by solving the pseudolikelihood equations, obtained by differentiating the log of the pseudolikelihood with respect to the elements of θ and setting them to zero; that is,

$$\frac{\partial}{\partial\theta_k}\log \mathrm{PL}(\delta,\theta)=\sum_{s\in S}\delta_s\sum_{i=1}^{N}\frac{\frac{\partial}{\partial\theta_k}f_s\left(x_i^{(s)};\theta\right)}{f_s\left(x_i^{(s)};\theta\right)}=0 \quad \text{for } k=1,\ldots,p. \tag{A2}$$

If regularity conditions given in AS for f and the f_s's are satisfied, then with probability tending to 1 as $N \to \infty$ the pseudolikelihood equation (A2) has a root $\hat{\theta}_N$ such that $\hat{\theta}_N \xrightarrow{N} \theta_0$, the true parameter value; i.e., the pseudolikelihood estimator is consistent. (The regularity conditions ensure, among other things, that the choice of δ does not omit any elements of a multidimensional θ from PL(δ,θ).) Moreover, the pseudolikelihood estimator is asymptotically normal. AS give an expression for its large-sample variance, which depends on the choice of δ and is bounded from below by the large-sample variance of the MLE. In the univariate case, any consistent sequence $\tilde{\theta}_N = \theta_N(X_1,\ldots,X_N)$ of roots of (A2) satisfies

$$\sqrt{N}\left(\tilde{\theta}_N - \theta_0\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{\mathrm{K}^\delta(\theta)}{\left[\mathrm{J}^\delta(\theta)\right]^2}\right), \tag{A3}$$

where

$$\mathrm{K}^\delta(\theta) = \sum_{s,s' \in S} \delta_s \delta_{s'} \mathrm{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log \mathrm{f}_s\left(X^{(s)};\theta\right)\right\}\left\{\frac{\partial}{\partial\theta}\log \mathrm{f}_{s'}\left(X^{(s')};\theta\right)\right\}\right]$$

and

$$\mathrm{J}^\delta(\theta) = -\sum_{s \in S} \delta_s \mathrm{E}_\theta\left\{\frac{\partial^2}{\partial\theta^2}\log \mathrm{f}_s\left(X^{(s)};\theta\right)\right\}.$$

Application to Expected Response Curves

The above results can be applied to the estimation of examinee ability under an IRT model. Let $\mathbf{X}=(X_1, \ldots, X_n)$ represent a response vector from an examinee to n items, governed by the IRT model $\mathrm{F}_j(\theta) \equiv \mathrm{P}\left(X_j = 1 \mid \theta, \beta_j\right)$ with

$$\mathrm{P}(\mathbf{X} = x \mid \theta, \mathbf{B}) = \prod_{j=1}^{N}\left[\mathrm{F}_j(\theta)\right]^{x_j}\left[1 - \mathrm{F}_j(\theta)\right]^{1-x_j}.$$

Let knowledge about **B** be expressed as p(**B**). The marginalized likelihood function for maximum likelihood estimation of θ is

$$\mathrm{P}(\mathbf{X} = x \mid \theta) = \int\left(\prod_{j=1}^{N}\left[\mathrm{F}_j(\theta)\right]^{x_j}\left[1 - \mathrm{F}_j(\theta)\right]^{1-x_j}\right)p(\mathbf{B})\partial\mathbf{B}.$$

For pseudolikelihood estimation, define δ as a selector for the subspace of S consisting of vectors that isolate a single item response; i.e.,

$$\delta_s = \begin{cases} 1 & \text{if } \sum_j s_j = 1 \\ 0 & \text{otherwise} \end{cases} .$$

The pseudolikelihood PL(δ,θ) corresponding to one observed response vector (i.e., N=1) is obtained by specializing (A1) as follows:

$$\begin{aligned} \mathrm{PL}(\delta,\theta) &= \prod_{s\in S}\left[\mathrm{f}_s\left(x^{(s)};\theta\right)\right]^{\delta_s} \\ &= \prod_{j=1}^{n}\mathrm{p}\left(x_j;\theta\right) \\ &= \prod_{j=1}^{n}\left[\mathrm{F}_j^*(\theta)\right]^{x_j}\left[1-\mathrm{F}_j^*(\theta)\right]^{1-x_j}, \end{aligned}$$

where $\mathrm{F}_j^*(\theta)$ is the expected response curve for Item j.

If knowledge about items is independent—i.e., p(**B**)=Πp(β_j)—then the asymptotic variance of the pseudolikelihood estimate (A3) simplifies to the usual inverse of the sum Fisher information over items, as calculated with expected response curves.

The AS consistency results imply the asymptotic equivalence of maximizing values of the full marginal likelihood, which does take dependencies among parameters from different items into account, and the product cf the expected response curves, which does not, for large samples of response vectors for the same θ. Since we typically observe only one response vector per examinee in practical work, small-sample behavior remains to be examined.

Appendix B

Program Documentation

This appendix provides detailed documentation for two computer programs: EXPRESFN and PLOTIRF. The EXPRESFN program computes EXPected RESponse FuNctions, both nonparametric and fitted, for a set of items, given a set of multivariate normal item parameter posterior distributions specified in terms of a set of mean vectors and an associated set of independent variance-covariance matrices. The PLOTIRF program provides plots of all estimated curves.

The EXPRESFN Program

The EXPRESFN program assumes that item responses may be modeled using a 2PL or a 3PL IRT model. Both nonparametric and fitted expected response functions are estimated for all items. The procedures used to estimate the fitted expected response functions are very similar to the procedures employed in LOGIST. The program also computes EAP ability estimates and standard errors for a set of examinees using the nonparametric and fitted expected response functions as well as the point estimates of the item parameter means.

The program has the following options:

1. The user may specify either a 2PL or a 3PL model.

2. The input point estimates of the item parameter means and variance-covariance matrices may be specified on the (a,b,c) scale or on the transformed (log(a),b,logit(c)) scale.

3. The range of the Θ grid and the total number of grid points may be specified.

4. In computing the fitted expected response function, the weighting distribution may be either normal or rectangular and the sum of the weights, ie. the total number of pseudo-examinees, may be specified.

5. In estimating the item parameters for the fitted expected response functions, the iterative procedure requires initial item parameter estimates. The program supplies default values for these initial estimates. However, the user may set all initial a's to a given value, all initial c's to a specified value or may supply the initial values.

6. To control the problem of estimating c's when the fitted expected response function becomes asymptotic below the minimum ability of interest, one may fix the c's at a common c for items where the estimated b-2/a is less than some criterion, fix all c's at a common c, put a beta prior on the c's and estimate the mean of the prior, or put a beta prior on the c's fixing the mean at a value specified by the user. The common c may be fixed or estimated.

7. Abilities may be estimated for an existing set of item responses or for a set of responses generated by the program for a random sample of examinees drawn from either a

normal or rectangular distribution. The generated data can be used to assess the differences between the abilities estimated using the three item response functions.

Nonparametric Expected Response Function

The nonparametric expected response function estimation procedure requires point estimates of the item parameters and associated variance-covariance matrices expressed on a transformed scale. If the input data has not already been transformed, then the following transformations will be applied:

$$a^t_j = \log(a_j)$$

$$b^t_j = b_j$$

$$c^t_j = \log(c_j/(1-c_j))$$

$$var(a^t_j) = var(a_j)/(a_j a_j)$$

$$cov(a^t_j,b^t_j) = cov(a_j,b_j)/a_j$$

$$cov(a^t_j,c^t_j) = cov(a_j,c_j)/(a_j c_j(1-c_j))$$

$$var(b^t_j) = var(b_j)$$

$$cov(b^t_j,c^t_j) = cov(b_j,c_j)/(c_j(1-c_j))$$

$$var(c^t_j) = var(c_j)/(c_j(1-c_j))^2$$

A grid of M Θ values are specified from Θ_{min} to Θ_{max}. Then a random sample of K parameter values are drawn from the multivariate normal distribution with means a^t_j, b^t_j, c^t_j and with the transformed variance-covariance matrix, $\Sigma^t(\beta_j)$. If the point estimate of c_j is 0, the c_j is held fixed and only log a_j and b_j sampled. The c_j for this item will also not be estimated for the fitted ERF. If the point estimate for c_j is less than or equal to .001, the mean for c_j used for the multivariate normal is set to the standard error of c. $F^*_j(\Theta_m)$ is computed for each of the M values of θ for each of the K IRF's. F^{**}_j is the average of the $F^*_j(\Theta_m)$'s.

Fitted Expected Response Function

The nonparametric ERF is the input for estimating the parameters of the fitted expected response function. The abilities are fixed at the Θ_m values in the grid. A sample of pseudo-examinees is generated to weight the grid values according to a weighting distribution specified by the user. The distribution may be either normal or rectangular. If normal the user may specify the mean and standard deviation. The user specifies the number of examinees for the sample. Newton's method is used to solve for the corrections to the estimated parameters by solving the likelihood equations. Since there are no omits, this procedure uses the expected values of the second derivatives which removes any possibilities of nonpositive definite matrices. If an item has a zero determinant, the item is removed from further estimation and the parameters are set to the values before the zero determinant.

The iteration procedure requires initial values for the item parameters. The default value for a is one. The default value for c is 1/(# choices) -.05. The default value for b is a function of the proportion correct. The formulas to compute the default values of b are:

$$b_j = h_j \frac{\sqrt{1+a_j^2}}{a_j}$$

where h_j is given by the following equations

$$p_j = \frac{1}{\sqrt{2\pi}} \int_{h_j}^{\infty} e^{-\frac{t^2}{2}} dt$$

and

$$p_j = \frac{\sum_{m=1}^{M} W_m F_j^*(\theta_m)}{N}$$

and N is the number of pseudo-examinees.

The procedure estimates the parameters for one item at a time until the relative change in a is less than .001 if a is being estimated. If a is fixed, the procedure iterates until the change in b is less than .001. One pass through all of the items constitutes a stage. In the first stage the c's are held fixed. In the second and following stages the c's are estimated unless a two parameter model is requested. If all c's are being estimated, or there is a prior on the c's, stages are repeated until the change in the likelihood is less than .02% between stages.

If no prior is imposed on the c's and the poorly estimated c's are restricted to a common c value, the following procedure is used:

In the second and third stages the c's for all items are estimated.

At the end of the third stage, the c's for items with b-2/a less than the criterion for fixing the c, (CRITFIXC), are fixed at a common c value. If all c's are to be fixed at a common c value, they are set to the common c value at this point.

The common c value is then estimated once per stage until the change in the common c is less than the standard error of the common c estimate for two successive stages. Only the items with c fixed at the common c are estimated in these stages.

The common c is then fixed and all items are again estimated until the criterion function increases by less than .02%

If a prior on c is requested and the mean is estimated, the mean is computed as the average of the c's at the end of each stage. Note: the beta prior is included in the computation of the likelihood and since the mean isn't actually a maximum likelihood estimate of the mean, the likelihood may not increase uniformly. To prevent premature stopping of the estimation procedure in this situation, the procedure will continue until the maximum difference between IRF's between stages is less than .001. The difference is computed for 5 abilities from -2 to 2 at intervals of 1.

The a parameter is restricted to a range of .01 to 99, c to a range of 0. to .99. The maximum amount that a parameter may change in any iteration is restricted. The amount for a is .1 times the previous value for a plus .2, b is .1 times previous value of b plus .4, and c is .06.

Input

The input to the program consists of a sysin file containing file names for the input and output files and parameters for controlling the procedure and a file containing the point estimates for the parameters and the variance-covariance of these estimates. If abilities are to be estimated for a group of examinees, the file of their responses is also read.

The Sysin File.

Record Set 1:

The first set of records in the sysin file define the input and output files. The set contains one record for each file to be defined. The last record in this set must be blank. The format for the file definition card is:

col 1	F
col 3 - 4	Unit number
col 6 - 45	File name, with all qualifiers

The files to be specified are:

Input files:

Unit 5	File containing the sysin dataset.
Unit 10	File containing, for each item, the point estimates and the variance-covariance matrix. They may be either on the a,b,c scale or on the log a, b, logit c scale but both the point estimates and the variance-covariance matrix must be on the same scale.
Unit 11	Input file containing the examinee responses if abilities are to be estimated for an existing item response file.

Output files:

Unit 6	Printed output file
Unit 7	Item parameter output in LOGIST7 format. The abilities written are the pseudo-abilities used to estimate the fitted ERF's.
Unit 12	Binary scratch output file, used to temporarily store the nonparametric ERF's and then the examinee responses.
Unit 13	Output file containing the point estimate item parameters, the fitted ERF, and the nonparametric ERF for each item.
Unit 14	Output file containing the sample of item response functions, if it was requested that the sample be saved.
Unit 15	Output file containing ability estimates, standard errors, and item responses, if abilities are estimated.

Record Set 2.

In record set 2, the options for running the procedure are specified. Only those options where the default says "Required" must be specified. The required parameters are the title, the number of items, the number of choices per item, and the format for reading the point estimates file. Defaults are supplied for all of the other parameters. The parameters are specified by entering the parameter name in positions 1 through 11 of the record and the value in positions 13 through 20. Formats are entered in positions 13 - 80. Right justify all integer values. The last record in this set must be blank.

Parameter input:

Parameter	Description / Options	Default
TITLE	Title for the run	required
#ITEMS	Number of items. (Maximum 800)	required
SEED	Random number seed. Integer between 0 and 1048576.	275927
DEBUG	Debugging printout?	NO
ITEMIDEN	Read in 8-character item identification codes?	NO

Parameter	Description / Options	Default
GENFIXC	Is c fixed in var/cov i.e. var/cov for c are 0? If so c will be fixed in fitting the ERF.	NO
IFTRANS	Are the input point estimates and var/cov matrix on the log a, b, logit c scale?	NO
FMTVAR	Format for reading point estimates and var/cov matrix. The values are read in following order: item number, a, b, c, var(a), cov(a,b), cov(a,c), var(b), cov(b,c), var(c). If abilities are to be estimated for a group of examinees, the item number must be the sequence number of the item in the record of item responses.	Required
#SAMPIRF	Number of item parameter values to sample (Maximum=1,000)	100
MINTHETA	Minimum ability for Θ grid	-3.
MAXTHETA	Maximum ability for Θ grid	3.
#ABILGRP	Number of points in Θ grid. (Maximum 201)	31
WEIGHTFN	Weighting distribution for fitting ERF. Enter RECTANGULAR or NORMAL	NORMAL
WEIGHTMN	If weighting distribution NORMAL, specify mean	0.
WEIGHTSD	If weighting distribution NORMAL, specify standard deviation.	1.
#ERFEXAM	Number of pseudo examinees for estimating the fitted ERF's. These will be apportioned by the weighting distribution to the M Θ grid points and adjusted so that there is an integral number of examinees at each grid point.	3100
SAVESAMP	Save the sample of item response functions to a file?	NO
READA	Read in initial a's?	NO
READB	Read in initial b's?	NO
READC	Read in initial c's?	NO

Parameter	Description / Options	Default
PRIORC	prior on c? 0 - no, estimate all c's, don't fix any at the common c value. 1 - no, fix c's at a common c (COMCx) if b-2/a<CRITFIXC. Estimate COMCx. 2 - no, fix all items at a common c. Estimate COMCx. 3 - yes, estimate the mean of prior. 4 - yes, fix the mean of prior.	0
CRITFIXC	Criterion for fixing c, if no prior requested and PRIORC = 1.	-2.5
AINIT	Initial a value, if READA is NO.	1.
AMAX	Maximum a.	99.0
PARMCODE	What parameters are to be estimated -1 - read in parmcode for each item Otherwise set parameter code for all items to the specified code. The definitions of the codes are: code parameters estimated 2 a,b 3 a,b,c	3
CHOICESx	Number of choices per item. x indicates a sequence number for different item types. Specify a different CHOICESx for each item type. For example, if a test has 4 and 5 choice items, set CHOICES1 to 4 and CHOICES2 to 5. x must be between 0 and 98.	Required
CINITx	Initial c for the CHOICESx items.	1/CHOICESx -.05
COMCx	If no prior on c , common c value for the CHOICESx items. If prior on c, mean c of prior for the CHOICESx items.	1/CHOICESx -.05

Parameter	Description / Options	Default
N-INFx	This is only used if there is a prior on c. It is the weight for the prior on c in terms of the number in a hypothetical group of examinees at minus infinity. It controls the variance of the beta prior. A separate N-INFx must be specified for every CHOICESx alternatives.	20
CHIx	Maximum c	.99
ESTABIL	Estimate abilities?	NO
#EXAMINEE	Number of examinees for which abilities are to be estimated if ESTABIL=YES. (Maximum 10,000)	20
PRIORMN	Prior mean of p(Θ)	0.
PRIORSD	Prior standard deviation of p(Θ)	1.
GENRESP	Generate artificial data, abilities and item responses.	YES
DISTABIL	If generating artificial data, specify type of ability distribution to generate, either 'RECTANGULAR' or 'NORMAL'.	RECTANGULAR
DISTMN	If DISTABIL is 'NORMAL', specify the mean of the distribution.	0.
DISTSD	If DISTABIL is 'NORMAL', specify the standard deviation of the distribution.	1.
RECTMIN	If DISTABIL is 'RECTANGULAR', specify minimum ability for distribution.	-3.
RECTMAX	If DISTABIL is 'RECTANGULAR', specify maximum ability for distribution.	3.
FMTRESP	If reading in examinee responses, specify format for reading the item responses. They will be selected as specified by item number read from the point estimates file. They are read in integer format. As many integer fields must be specified as the maximum item number read from the point estimates. For example, if the item numbers read from the point estimates are 1,5, and 10. The format must specify reading in 10 integer fields.	Required if ESTABIL=YES and GENRESP=NO.

Additional input:

If PARMCODE = -1, read in a parameter code for each item with Record set 3.
If more than one CHOICESx read, specify the items for each number of choices in Record set 4.
If ITEMIDEN requested, read in item identification in Record set 5.

Record set 3.

This record set is only required if PARMCODE is set to -1 to read in a parameter code for each item.

col 1 - 8 "PARMCODE"
col 9-10 Sequence number for this PARMCODE record.
col 11-80 Parameter codes for the items in 35I2 format.

Repeat for as many records as necessary, increasing the sequence number for each record. For example, for items 36-40, the sequence number must be 2.

Record set 4.

This record set is only necessary if more than one CHOICESx is specified. It is used to specify the number of choices for each item.

col 1 - 8 "CHOICESx" where x corresponds to the CHOICESx specified on the parameter records.
col 9 -10 Sequence number for this CHOICESx record.
col 11 - 80 Item numbers of the items, that have the number of choices specified by CHOICESx, read in (10I5) format. A sequence of items can be specified by specifying the first number in the sequence followed by the negative of the last number in the sequence.

Enter as many CHOICESx records as necessary, increasing the sequence number for each record. Do no split a sequence across two records. If the beginning of a sequence would be the last field of a record, leave the last field blank and start the sequence on the next record.

Record set 5.

If ITEMIDEN is "YES", this set is required to read in the 8-character item identification for each item.

col 1 - 8 "ITEMIDEN"
col 9 - 10 Sequence number
col 11 - 18 Item identification for the first item. Left justify the identification in the field.
col 19 - 10 Blank
col 21 - 28 Item identification for the second item.
col 29 - 30 Blank
etc. etc.

Enter 7 item identifications per record, repeat for as many records as necessary, increasing the sequence number for each record. For example, record with sequence number 2 will contain the identifications for items 8 through 14.

<u>Detailed description of output:</u>

Unit 6 Printed output file

The printout contains:

Check on input parameters and defaults.

For the nonparametric ERF, the point estimates, the input var/cov matrix, the var/cov for the sampled IRF's for both the a,b,c scale and the transformed scale, and the nonparametric ERF for a spaced sample of the Θ grid points are printed.

For the estimation of the parameters for the fitted ERF, the likelihood is printed for each stage as well as the maximum derivatives for the three parameters, the maximum change in an iteration, and the maximum change over all iterations for each type of parameter. If the common c is being computed, information on the computation of the common c values is printed.

For each item there is a parameter code that indicates which item parameters are being estimated. The values for the codes are defined in the input description. In addition, a 20 is added to the code if the c for an item is held fixed at the common c. If an item is removed because the expected matrix of second derivatives had a zero determinant, the parameter code is set to 996.

The final item parameter estimates are printed as well as the standard errors of the estimates.

If abilities are estimated, the EAP ability estimates and the standard errors are printed for the point estimate IRF, the nonparametric ERF and the fitted ERF. Only the first and last 10 are printed.

Unit 7 Item parameter output in LOGIST7 format. The abilities written are the pseudo-abilities used to estimate the fitted ERF's. A subroutine to read this file is included with the program. The subroutine contains comment statements that describe the calling arguments. Output includes the title, the number of items, the number of pseudo-examinees, the estimated item parameters, the pseudo-abilities, variables used in the estimation of c, and parameter code indicator for number of parameters estimated.

Unit 13 File containing the nonparametric item response functions for plotting with the plot program. The first record contains the title of the run. The second record contains the number of items (I5). The third record contains the M abilities for the Θ grid in the format (5X,10F8.4). The remaining records contain the item sequence number, the item number, the item identification, the a,b,c point estimates, a,b,c estimates for the fitted ERF, the parameter code,and the nonparametric proportion correct for the M abilities in the format (2I5,A8,1X,3F12.6,1X,3F12.6,I4/(10F12.6))

Unit 14 Output file containing the sample of item response functions, if it was requested that it be saved. For each item, the item number and the three parameters for each sampled IRF are written in the format (I4,12F12.6/(4X,12F12.6)).

Record 1: col 1 - 4: Item number
col 5 - 16: a for first item sampled
col 17 -28: b for first item sampled
col 29 - 40: c for first item sampled
col 41 - 52: a for second item sampled
etc. etc.

Unit 15 Output file containing ability estimates and standard errors, and item responses, if abilities are estimated.

For each examinee a record is written in the format (I5,7F12.6,600I1) containing:

col 1 - 5: examinee sequence number
col 6 - 17 - true ability, (if responses are read, this is set to 999999.)
col 18 - 29 - EAP ability computed using point estimate IRF
col 30 - 41 - EAP ability computed using fitted ERF
col 42 - 53 - EAP ability computed using nonparametric ERF
col 54 - 65 - Standard error of ability computed using point estimate IRF
col 66 - 77 - Standard error of ability computed using the fitted ERF
col 78 - 89 - Standard error of ability computed using nonparametric ERF
col 90 + Item responses in I1 format, items 1 to #ITEMS.

The PLOTIRF Program

A plot program was also developed that plots the three item response functions for comparison of the three curves. This program produces plots on the screen, a laser printer, or a postscript printer. Input to the program consists of a sysin file with the control parameters and the file written on the unit 13 by the EXPRESFN program. One, four or eight plots per page are possible.

Input

The sysin file consists of a set of records defining the input and output files and a few control parameters.

Record set defining files.
The set contains one record for each file to be defined.
The last record in this set must be blank.

The format for the file definition card is:

col 1	F
col 3 - 4	Unit number
col 6 - 45	File name, with all qualifiers

The files to be specified are:

Input files:

Unit 5	Sysin file containing file definitions and parameters.
Unit 13	File written on unit 13 in EXPRESFN containing the nonparametric item response functions.

Output file:

Unit 9	Plot output if requested that the plots be saved for printing later.

Record set specifying control parameters.
The last record in this set must be a blank record.

Parameter	Description/options	Default
TITLE	Title for plots.	Title from EXPRESFN.
IFSELIT	Select items from items in EXPRESFN run.	NO
PLOTDEV	Plotting device: POSTSCRIPT LASER - HP laser printer SCREEN - only display on screen.	LASER
#PLOTPAGE	Number of plots per page. Options are 1, 4, or 8.	8
SAVEPLOT	Plot now or write plots to file? NO - print plots now YES - save plots to a file for printing later.	NO

Record set 3.

If IFSELIT is YES to select items from the EXPRESFN run, specify the items to select with this record set.

The format of record set 3 is as follows:

col 1 - 8	"IFSELIT"
col 9 -10	Sequence number for this IFSELIT record.
col 11 - 15	Item number of first item to be selected.
col 16 - 20	Item number of second item to be selected.
etc.	etc.
col 76 - 80	Item number of 14th item to be selected.

Indicate a sequence of item numbers by entering the first in the sequence and the negative of the last in the sequence. Repeat for as many cards as necessary. Increase the sequence number for each card. Do not split a sequence across two records. If the beginning of a sequence would be the last field of a record, leave the last field blank and start the sequence on the next record.

Distribution List

Dr Terry Ackerman
Educational Psychology
260C Education Bldg
University of Illinois
Champaign IL 61801

Dr Terry Allard
Code 3422
Office of Naval Research
800 N Quincy St
Arlington VA 22217-5660

Dr Nancy Allen
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Gregory Anrig
Educational Testing Service
Mail Stop 14-C
Princeton NJ 08541

Dr Phipps Arabie
Graduate School of Management
Rutgers University
92 New Street
Newark NJ 07102-1895

Dr Isaac I Bejar
Educational Testing Service
Mail Stop 11-R
Princeton NJ 08541

Dr William O Berry
Director
Life and Environmental Sciences
AFOSR/NL N1
Bldg 410
Bolling AFB DC 20332-6448

Dr Thomas G Bever
Department of Psychology
University of Rochester
River Station
Rochester NY 14627

Dr Menucha Birenbaum
School of Education
Tel Aviv University
Ramat-Aviv 69978 ISRAEL

Dr Bruce Bloxom
Defense Manpower Data Center
99 Pacific St
Suite 155A
Monterey CA 93943-3231

Dr Gwyneth Boodoo
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Richard L Branch
HQ USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago IL 60064

Dr Robert Brennan
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr David V Budescu
Department of Psychology
University of Haifa
Mount Carmel Haifa 31999
ISRAEL

Dr Gregory Candell
CTB/MacMillan/McGraw-Hill
2500 Garden Road
Monterey CA 93940

Dr Paul R Chatelier
PERCEPTRONICS
1911 North Ft Myer Drive
Suite 1100
Arlington VA 22209

Dr Susan Chipman
Cognitive Science Program
Office of Naval Research
800 North Quincy Street
Code 3422
Arlington VA 22217-5660

Dr Raymond E Christal
UES LAMP Science Advisor
AL/HRMIL
Brooks AFB TX 78235

Dr Norman Cliff
Department of Psychology
University of Southern California
Los Angeles CA 90089-1061

Director
Life Sciences
Code 3420
Office of Naval Research
Arlington VA 22217-5660

Commanding Officer
Naval Research Laboratory
Code 4827
Washington DC 20375-5000

Dr John M Cornwell
Department of Psychology
I/O Psychology Program
Tulane University
New Orleans LA 70118

Dr William Crano
Department of Psychology
Texas A&M University
College Station TX 77843

Dr Linda Curran
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn VA 22209

Professor Clément Dassa
Faculté des sciences de l'éducation
Département d'études en éducation
et d'administration de l'éducation
CP 6128 succursale A
Montéal Québec
CANADA H3C 3J7

Dr Timothy Davey
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Charles E Davis
Educational Testing Service
Mail Stop 16-T
Princeton NJ 08541

Dr Ralph J DeAyala
Meas Stat and Eval
Benjamin Bldg Room 1230F
University of Maryland
College Park MD 20742

Dr Sharon Derry
Florida State University
Department of Psychology
Tallahassee FL 32306

Hei-Ki Dong
BELLCORE
6 Corporate Place
RM: PYA-1K207
PO Box 1320
Piscataway NJ 08855-1320

Dr Neil Dorans
Educational Testing Service
Mail Stop 07-E
Princeton NJ 08541

Dr Fritz Drasgow
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Defense Tech Information Center
Cameron Station Bldg 5
Alexandria VA 22314
(2 Copies)

Dr Richard Duran
Graduate School of Education
University of California
Santa Barbara CA 93106

Dr Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence KS 66045

Dr George Engelhard Jr
Division of Educational Studies
Emory University
210 Fishburne Bldg
Atlanta GA 30322

ERIC Facility-Acquisitions
2440 Research Blvd
Suite 550
Rockville MD 20850-3238

Dr Marshall J Farr
Farr-Sight Co
2520 North Vernon Street
Arlington VA 22207

Dr Leonard Feldt
Lindquist Center for Measurement
University of Iowa
Iowa City IA 52242

Dr Richard L Ferguson
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Gerhard Fischer
Liebiggasse 5
A 1010 Vienna
AUSTRIA

Dr Myron Fischl
US Army Headquarters
DAPE-HR
The Pentagon
Washington DC 20310-0300

Mr Paul Foley
Navy Personnel R&D Center
San Diego CA 92152-6800

Chair
Department of Computer Science
George Mason University
Fairfax VA 22030

Dr Robert D Gibbons
University of Illinois at Chicago
NPI 909A M/C 913
912 South Wood Street
Chicago IL 60612

Dr Janice Gifford
University of Massachusetts
School of Education
Amherst MA 01003

Dr Robert Glaser
Learning Res & Development Cntr
University of Pittsburgh
3939 O'Hara Street
Pittsburgh PA 15260

Dr Susan R Goldman
Peabody College
Box 45
Vanderbilt University
Nashville TN 37203

Dr Timothy Goldsmith
Department of Psychology
University of New Mexico
Albuquerque NM 87131

Dr Sherrie Gott
AFHRL/MOMJ
Brooks AFB TX 78235-5601

Dr Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore MD 21218

Professor Edward Haertel
School of Education
Stanford University
Stanford CA 94305-3096

Dr Ronald K Hambleton
University of Massachusetts
Lab of Psychom & Eval Res
Hills South Room 152
Amherst MA 01003

Dr Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign IL 61820

Dr Patrick R Harrison
Computer Science Department
US Naval Academy
Annapolis MD 21402-5002

Ms Rebecca Hetter
Navy Personnel R&D Center
Code 13
San Diego CA 92152-6800

Dr Thomas M Hirsch
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Professor Paul W Holland
Div of Educ Psych & Quant
Methods Prog
Graduate School of Education
4511 Tolman Hall
University of California-Berkeley
Berkeley CA 94720

Professor Lutz F Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Ms Julia S Hough
Cambridge University Press
40 West 20th Street
New York NY 10011

Dr William Howell
Chief Scientist
AFHRL/CA
Brooks AFB TX 78235-5601

Dr Huynh Huynh
College of Education
University of South Carolina
Columbia SC 29208

Dr Martin J Ippel
Center for the Study of
Education and Instruction
Leiden University
PO Box 9555
2300 RB Leiden
THE NETHERLANDS

Dr Robert Jannarone
Elec and Computer Eng Dept
University of South Carolina
Columbia SC 29208

Dr Kumar Joag-dev
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Professor Douglas H Jones
Grad Sch of Management
Rutgers The State University NJ
Newark NJ 07102

Dr Brian Junker
Carnegie-Mellon University
Department of Statistics
Pittsburgh PA 15213

Dr Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh PA 15213

Dr J L Kaiwi
Code 442/JK
Naval Ocean Systems Center
San Diego CA 92152-5000

Dr Michael Kaplan
Office of Basic Research
US Army Research Institute
5001 Eisenhower Avenue
Alexandria VA 22333-5600

Dr Jeremy Kilpatrick
Dept of Mathematics Education
105 Aderhold Hall
University of Georgia
Athens GA 30602

Ms Hae-Rim Kim
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr. Jwa-keun Kim
Department of Psychology
Middle Tennessee State University
Murfreesboro TN 37132

Dr Sung-Hoon Kim
KEDI
92-6 Umyeon-Dong
Seocho-Gu
Seoul
SOUTH KOREA

Dr G Gage Kingsbury
Portland Public Schools
Res & Eval Department
501 North Dixon Street
PO Box 3107
Portland OR 97209-3107

Dr William Koch
Box 7246
Meas & Eval Center
University of Texas-Austin
Austin TX 78703

Dr James Kraatz
Computer-based Education
Research Laboratory
University of Illinois
Urbana IL 61801

Dr Patrick Kyllonen
AFHRL/MOEL
Brooks AFB TX 78235

Ms Carolyn Laney
1515 Spencerville Rod
Spencerville MD 20868

Richard Lanterman
Commandant (G-PWP)
US Coast Guard
2100 Second Street SW
Washington DC 20593-0001

Dr Michael Levine
Educational Psychology
210 Education Building
1310 South Sixth Street
Univ of IL at Urbana-Champaign
Champaign IL 61820-6990

Dr Charles Lewis
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541-0001

Mr Hsin-hung Li
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Library
Naval Training Systems Center
12350 Research Parkway
Orlando FL 32826-3224

Dr Marcia C Linn
Graduate School of Education
EMST
Tolman Hall
University of California
Berkeley CA 94720

Dr Robert L Linn
Campus Box 249
University of Colorado
Boulder CO 80309-0249

Logicon Inc (Attn: Library)
Tactical & Training Systems Div
PO Box 85158
San Diego CA 92138-5158

Dr Richard Luecht
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr George B. Macready
Dept of Meas Stat & Eval
College of Education
University of Maryland
College Park MD 20742

Dr Evans Mandes
George Mason University
4400 University Drive
Fairfax VA 22030

Dr Paul Mayberry
Center for Naval Analysis
4401 Ford Avenue
PO Box 16268
Alexandria VA 22302-0268

Dr James R McBride
HumRRO
6430 Elmhurst Drive
San Diego CA 92120

Mr Christopher McCusker
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Dr Joseph McLachlan
Navy Pers Res & Dev Cntr
Code 14
San Diego CA 92152-6800

Alan Mead
c/o Dr Michael Levine
Educational Psychology
210 Education Bldg
University of Illinois
Champaign IL 61801

Dr Timothy Miller
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Robert Mislevy
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Ivo Molenar
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
The NETHERLANDS

Dr Eiji Muraki
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Ratna Nandakumar
Educational Studies
Willard Hall Room 213E
University of Delaware
Newark DE 19716

Acad Prog & Research Branch
Naval Tech Training Command
Code N-62
NAS Memphis (75)
Millington TN 30854

Dr W Alan Nicewander
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Head
Personnel Systems Department
NPRDC (Code 12)
San Diego CA 92152-6800

Director
Training Systems Department
NPRDC (Code 14)
San Diego CA 92152-6800

Library NPRDC
Code 041
San Diego CA 92152-6800

Librarian
Naval Cntr for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington DC 20375-5000

Office of Naval Research
Code 3422
800 N Quincy Street
Arlington VA 22217-5660
(6 Copies)

ONR Resident Representative
New York City
33 Third Avenue - Lower Level
New York NY 10003-9998

Special Asst for Res Management
Chief of Naval Personnel
(PERS-O1JT)
Department of the Navy
Washington DC 20350-2000

Dr Judith Orasanu
NASA Ames Research Center
Mail Stop 239-1
Moffett Field CA 94035

Dr Peter J Pashley
Law School Admission Services
PO Box 40
Newtown PA 18940-0040

Wayne M Patience
American Council on Education
GED Testing Service Suite 20
One Dupont Circle NW
Washington DC 20036

Dept of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey CA 93943-5026

Dr Peter Pirolli
School of Education
University of California
Berkeley CA 94720

Dr Mark D Reckase
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Mr Steve Reise
Department of Psychology
University of California
Riverside CA 92521

Mr Louis Roussos
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr Donald Rubin
Statistics Department
Science Center Room 608
1 Oxford Street
Harvard University
Cambridge MA 02138

Dr Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg
Knoxville TN 37966-0900

Dr Mary Schratz
4100 Parkside
Carlsbad CA 92008

Mr Robert Semmes
N218 Elliott Hall
Department of Psychology
University of Minnesota
Minneapolis MN 55455-0344

Dr Valerie L Shalin
Dept of Industrial Engineering
State University of New York
342 Lawrence D Bell Hall
Buffalo NY 14260

Mr Richard J Shavelson
Graduate School of Education
University of California
Santa Barbara CA 93106

Kathleen Sheehan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr Randall Shumaker
Naval Research Laboratory
Code 5500
4555 Overlook Avenue SW
Washington DC 20375-5000

Dr Judy Spray
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Martha Stocking
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St
Champaign IL 61820

Dr Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr David Thissen
Psychometric Laboratory
CB# 3270 Davie Hall
University of North Carolina
Chapel Hill NC 27599-3270

Mr Thomas J Thomas
Federal Express Corporation
Human Resource Development
3035 Director Row Suite 501
Memphis TN 38131

Mr Gary Thomasson
University of Illinois
Educational Psychology
Champaign IL 61820

Dr Howard Wainer
Educational Testing Service
15-T Rosedale Road
Princeton NJ 08541

Elizabeth Wald
Office of Naval Technology
Code 227
800 North Quincy Street
Arlington VA 22217-5000

Dr Michael T Waller
Univ of Wisconsin-Milwaukee
Educ Psychology Department
Box 413
Milwaukee WI 53201

Dr Ming-Mei Wang
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Thomas A Warm
FAA Academy
PO Box 25082
Oklahoma City OK 73125

Dr David J Weiss
N660 Elliott Hall
University of Minnesota
75 E River Road
Minneapolis MN 55455-0344

Dr Douglas Wetzel
Code 15
Navy Personnel R&D Center
San Diego CA 92152-6800

German Military Representative
Personalstammamt
Koelner Str 262
D-5000 Koeln 90
WEST GERMANY

Dr David Wiley
Sch of Educ and Social Policy
Northwestern University
Evanston IL 60208

Dr Bruce Williams
Dept of Educational Psychology
University of Illinois
Urbana IL 61801

Dr Mark Wilson
School of Education
University of California
Berkeley CA 94720

Dr Eugene Winograd
Department of Psychology
Emory University
Atlanta GA 30322

Dr Martin F Wiskoff
PERSEREC
99 Pacific Street
Suite 4556
Monterey CA 93940

Mr John H Wolfe
Navy Personnel R&D Center
San Diego CA 92152-6800

Dr Kentaro Yamamoto
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Duanli Yan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey CA 93940

Dr Joseph L Young
National Science Foundation
Room 320
1800 G Street NW
Washington DC 20550